

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-028391

(43)Date of publication of application : 04.02.1994

(51)Int.Cl.

G06F 15/38
G06F 15/20
G06F 15/20
G06F 15/40

(21)Application number : 04-202971

(71)Applicant : HITACHI LTD

(22)Date of filing : 07.07.1992

(72)Inventor : KIYAMA TADAHIRO
KINUKAWA HIROYUKI
TSUJI HIROSHI

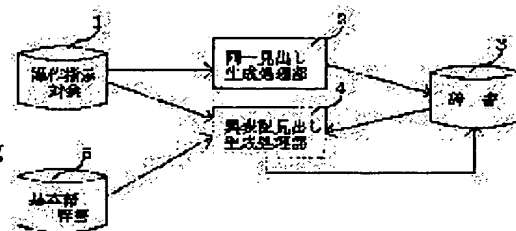
(54) DICTIONARY GENERATION SYSTEM

(57)Abstract:

PURPOSE: To provide the system for generating a dictionary by generating word information taking character strings constituting the word in a database as heading character strings and registering them in a batch.

CONSTITUTION: A word in the database being an object 1 of an operation instruction is acquired by a same entry generation processing section 2. Character strings are generated based on the word to make it the heading character string of dictionary information storing the word information, and a dictionary 3 having word information having correspondence relation with the words in the database is generated. Based on the generated dictionary, a different description heading generation processing section 4 takes character strings which can be obtained by replacing the full size character of the heading character string with a half size character, the half size character with the full size character, characters written in the HIRAGANA (Japanese syllabary) system with characters written in the KATAKANA (the square form of the Japanese syllabary) system, and

characters written in the KATAKANA system with the characters written in the HIRAGANA system, partial character strings common to the heading character strings, and character strings which can be obtained by arranging the group of head characters and the group of end characters of the partial character string in the heading character string as new heading character strings, and generating the dictionary 3.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平6-28391

(43)公開日 平成6年(1994)2月4日

(51)Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/38	C	9194-5L		
15/20	5 2 0 N	6798-5L		
	L	6798-5L		
	5 2 2 K	6798-5L		
15/40	5 0 0 S	7218-5L		

審査請求 未請求 請求項の数6(全19頁)

(21)出願番号 特願平4-202971

(22)出願日 平成4年(1992)7月7日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 木山 忠博

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72)発明者 絹川 博之

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72)発明者 辻 洋

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(74)代理人 弁理士 笹岡 茂 (外1名)

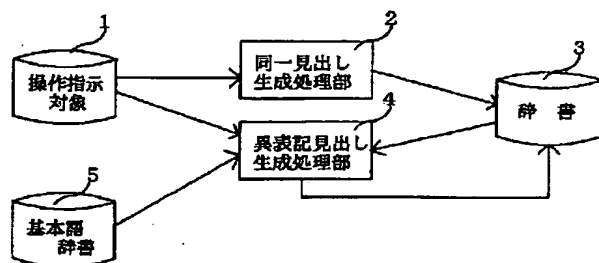
(54)【発明の名称】 辞書生成システム

(57)【要約】

【目的】 データベース内の単語を構成する文字列を見出し文字列とした単語情報を生成し、これを一括して登録して辞書を生成するシステムを提供する。

【構成】 同一見出し生成処理部2により、操作指示の対象1であるデータベース内の語を取得し、この語を基に文字列を生成して単語情報を格納する辞書情報の見出し文字列としデータベース中の単語との対応関係を持つ単語情報を有する辞書3を生成する。また、この生成した辞書を基にして、異表記見出し生成処理部4により、前記見出し文字列の全角文字を半角文字に、半角文字を全角文字に、平仮名文字を片仮名文字に、片仮名文字を平仮名文字にそれぞれ置換して得られる文字列、また、前記見出し文字列同士で共通な部分文字列、さらに見出し文字列内の部分文字列の先頭文字同士、末尾文字同士を並べて得られた文字列を新たな見出し文字列として辞書3を生成する。

【図1】



【特許請求の範囲】

【請求項1】 入出力装置と、処理装置と、操作指示対象が格納されかつ辞書を格納するエリアを有する記憶装置を備え、操作指示対象を基に辞書を生成する辞書生成システムであって、

前記処理装置は前記記憶装置に格納された操作指示対象内の各単語を構成する文字列を順次読み出し該文字列を辞書の見出し文字列として前記記憶装置の辞書を格納するエリアに順次格納し、見出し文字列に付随する各項目の項目データを生成して前記記憶装置の辞書を格納するエリアに前記各見出し文字列対応に順次格納して辞書を生成する手段を備えることを特徴とする辞書生成システム。

【請求項2】 入出力装置と、処理装置と、基本語辞書および請求項1記載の辞書生成システムで生成された辞書が格納されかつ辞書を格納するエリアを有する記憶装置を備え、前記生成された辞書を基に新たな辞書を生成する辞書生成システムであって、

前記処理装置は、
前記生成された辞書の各見出し文字列を解析し、見出し文字列を構成する文字のバイトコードが1バイトコードか2バイトコードであるかを判定する手段と、
判定されたバイトコードを他のバイトコードに置換可能か否か判定する手段と、置換可能と判定されたとき、
前記見出し文字列の内の置換可能な文字列を他のバイトコードに置換し、得られた見出し文字列を新たな見出し文字列とし、元の見出し文字列に付随する項目データを新たな見出し文字列に付随する項目データとして新たな辞書を生成する手段を備えることを特徴とする辞書生成システム。

【請求項3】 入出力装置と、処理装置と、基本語辞書および請求項1記載の辞書生成システムで生成された辞書が格納されかつ辞書を格納するエリアを有する記憶装置を備え、前記生成された辞書を基に新たな辞書を生成する辞書生成システムであって、

前記処理装置は、
前記生成された辞書の各見出し文字列を解析し、見出し文字列中に平仮名文字あるいは片仮名文字が含まれているか判定する手段と、

判定された平仮名文字を片仮名文字に、片仮名文字を平仮名文字に置換し、得られた見出し文字列を新たな見出し文字列とし、元の見出し文字列に付随する項目データを新たな見出し文字列に付随する項目データとして新たな辞書を生成する手段を備えることを特徴とする辞書生成システム。

【請求項4】 入出力装置と、処理装置と、基本語辞書および請求項1記載の辞書生成システムで生成された辞書が格納されかつ辞書を格納するエリアを有する記憶装置を備え、前記基本語辞書および生成された辞書を基に新たな辞書を生成する辞書生成システムであって、

前記処理装置は、

前記基本語辞書を参照して前記生成された辞書の各見出し文字列を解析し、各見出し文字列毎に形態素解析結果レコードを得る手段と、

得られた1つの形態素解析結果レコードから単語分割結果を取り出し、該単語分割結果の中の分割された文字列を検索キーとして該検索キーと等しい文字列を有する他の形態素解析結果レコードを検索し、他の形態素解析結果レコードが検索されたとき、前記検索キーとなった分割された文字列を新たな見出し文字列とし、検索された他の形態素解析結果レコードの元となった見出し文字列に付随する項目データを新たな見出し文字列に付随する項目データとし、各形態素解析結果レコード対応に得られた新たな見出し文字列とその項目データから新たな辞書を生成する手段を備えることを特徴とする辞書生成システム。

【請求項5】 入出力装置と、処理装置と、基本語辞書および請求項1記載の辞書生成システムで生成された辞書が格納されかつ辞書を格納するエリアを有する記憶装置を備え、前記基本語辞書および生成された辞書を基に新たな辞書を生成する辞書生成システムであって、

前記処理装置は、
前記基本語辞書を参照して前記生成された辞書の各見出し文字列を解析し、各見出し文字列毎に形態素解析結果レコードを得る手段と、

得られた形態素解析結果レコードから単語分割結果を取り出し、該単語分割結果の中の各分割された文字列から先頭文字1文字を順次取り出し、順次取り出された文字を並べて得られる文字列を新たな見出し文字列とし、元の見出し文字列に付随する項目データを新たな見出し文字列に付随する項目データとし、各形態素解析結果レコード対応に得られた新たな見出し文字列とその項目データから新たな辞書を生成する手段を備えることを特徴とする辞書生成システム。

【請求項6】 入出力装置と、処理装置と、基本語辞書および請求項1記載の辞書生成システムで生成された辞書が格納されかつ辞書を格納するエリアを有する記憶装置を備え、前記基本語辞書および生成された辞書を基に新たな辞書を生成する辞書生成システムであって、

前記処理装置は、
前記基本語辞書を参照して前記生成された辞書の各見出し文字列を解析し、各見出し文字列毎に形態素解析結果レコードを得る手段と、

得られた形態素解析結果レコードから単語分割結果を取り出し、該単語分割結果の中の各分割された文字列から末尾文字1文字を順次取り出し、順次取り出された文字を並べて得られる文字列を新たな見出し文字列とし、元の見出し文字列に付随する項目データを新たな見出し文字列に付随する項目データとし、各形態素解析結果レコード対応に得られた新たな見出し文字列とその項目データ

3

タから新たな辞書を生成する手段を備えることを特徴とする辞書生成システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、日本語による情報検索等の処理を行う情報処理システムにおける日本語解析で用いる辞書生成システムに係り、特に、操作指示対象であるデータベース中の単語の単語情報を基に辞書情報生成し格納して好適な辞書を生成する辞書生成システムに関する。

【0002】

【従来の技術】従来の、辞書作成システムは、例えば、特開平 2 - 1 6 3 8 7 4 号公報に記載されている。これは、辞書への単語登録時間の遅延を防ぐことを目的としたものである。具体的には、オペレータにより確認された文字列を文字コード種別に基づき単語分割し、単語分割の結果を利用して単語候補を抽出し、この単語が辞書に登録されているか否か判別し、未登録である場合に該当する単語を辞書に登録するものである。

【0003】

【発明が解決しようとする課題】従来の、計算機における辞書作成システムでは、利用者が確認し指定した単語に対する辞書登録に留まっている。例えば、特開平 2 - 1 6 3 8 7 4 号公報に記載されている単語辞書作成方式は、入力文字列を文字コード種別に基づいて単語境界を検出し、1 個以上の単語領域に分割し単語候補を抽出し、辞書に未登録の単語候補が存在するならば、該単語候補が単語として成立するか否か判別し、成立する場合に単語辞書に追加登録するものである。この従来方式は自然語解析中に辞書登録すべき単語を検出し逐次的に辞書登録するものであり、自然語解析が対象とする世界の情報を持つ単語情報を一括して辞書登録するものではない。本発明の目的は、自然語解析が対象とする世界の情報を持つ単語情報を一括して辞書登録し、さらに、対象世界に存在する単語の同一表記だけではなく、利用者が容易に類推可能な単語の表記を生成する辞書生成システムを提供することにある。

【0004】

【課題を解決するための手段】上記目的を達成するために、本発明の辞書生成システムは、操作指示の対象であるデータベース内の語を取得し、この語を基に文字列を生成して単語情報を格納する辞書情報の見出し文字列としデータベース中の単語との対応関係を持つ単語情報を有する辞書を生成する。辞書の見出し文字列としては、データベース内の単語を構成する文字列を見出し文字列とする。

【0005】さらに、前記辞書の見出し文字列を構成する文字のバイトコードが 1 バイトコードか 2 バイトコードであるかを判定して、判定されたバイトコードを他のバイトコードに置換できるときには置換可能な文字列を

4

他のバイトコードに置換して新たな文字列とし、この文字列を新たな見出し文字列とする新たな辞書を生成する。さらに、前記辞書の見出し文字列中に平仮名文字あるいは片仮名文字が含まれているか判定し、判定された平仮名文字を片仮名文字に、片仮名文字を平仮名文字に置換し、得られた文字列を新たな見出し文字列とする新たな辞書を生成する。さらに、基本語辞書を参照して前記辞書の見出し文字列を解析し、各見出し文字列毎に形態素解析結果レコードを得、得られた 1 つの形態素解析結果レコードの単語分割結果の中の分割された文字列を検索キーとして該検索キーと等しい文字列を有する他の形態素解析結果レコードを検索し、他の形態素解析結果レコードが検索されたとき、前記検索キーとなった分割された文字列を新たな見出し文字列とする新たな辞書を生成する。さらに、前記の各形態素解析結果レコードから単語分割結果を取り出し、該単語分割結果の中の各分割された文字列から先頭文字 1 文字を順次取り出し順次取り出された文字を並べて得られる文字列を新たな見出し文字列とする新たな辞書を生成する。また、各分割された文字列から末尾文字 1 文字を順次取り出し順次取り出された文字を並べて得られる文字列を新たな見出し文字列とする新たな辞書を生成する。

【0006】

【作用】本発明により、自然語解析が対象とする操作指示対象の情報を持つ単語情報を一括して辞書登録することが可能になる。さらに、操作指示対象内に存在する単語と同一表記（見出し）文字列を持つ単語情報のみではなく、利用者が容易に類推可能であり使用すると予測される同類語や略語等の単語の表記文字列を持つ単語情報を生成し辞書登録することが可能になり、辞書情報を充実することができる。また、本発明により、単語の辞書登録に要する非常に膨大な作業を省くことができる。

【0007】

【実施例】以下本発明の実施例を、図面により詳細に説明する。図 1 は、本発明の辞書生成システムの一実施例を示す機能ブロック図である。同一見出し生成処理部 2 は、操作指示対象 1 を参照し、操作指示対象内の語と等しい文字列を見出し文字とする単語を生成し、辞書 3 に登録する。異表記見出し生成処理部 4 は、操作指示対象 1 と基本語辞書 5 を参照し、操作指示対象内の語と異なる文字列を見出し文字とする単語、即ち同類語を生成し、辞書 3 に登録する。このように単語の生成機能を持たせることにより、辞書の自動生成を実現する。

【0008】図から明らかなように、同一見出し生成処理部 2、異表記見出し生成処理部 4 は処理装置で行なわれる処理を示し、操作指示対象（データベース）1、基本語辞書 5、辞書 3 はファイル（テーブルとも呼ぶ）である。このように、本実施例によれば、各機能ブロックが、プログラム論理により構成されている。そのため、各機能ブロック単位に L S I 化が可能であり、知識獲得

5

装置として、処理の高速化を図ることができる。

【0009】図2は、図1における辞書生成システムの全体的なハードウェア構成を示すブロック図である。入出力装置6は、データの入力を行ない、そして各種情報を表示するためのものである。プロセッサ7は、プログラムに基づき、図1における処理を実行する。記憶装置8は、図1における基本語辞書5や各種プログラムを格納する。さらに、記憶装置8は、プロセッサ7の各処理実行用のメモリであるワーキングエリア9、同一見出し生成処理部格納エリア20、異表記見出し生成処理部格納エリア40、辞書格納エリア30、基本語辞書格納エリア40、操作指示対象格納エリア10の記憶部を持っている。記憶装置8に格納される各プログラムは、プロセッサ7において実行される。その実行に際して、必要に応じて入出力装置6が用いられる。

【0010】図3は、図1における同一見出し生成処理のPAD図(Problem Analysis Diagram)である。操作指示対象1であるデータベース内の語の文字列を見出し文字列とする単語情報とデータベース内の語との対応関係を持つ単語情報を生成し、辞書3に登録するまでの処理を示したものである。図7は、図1の操作指示対象1であるデータベースの例であり、「商店テーブル」と「商品テーブル」の2つのテーブルにより構成されており、「商品テーブル」は「商品名」、「価格」、「店名」の3つのカラム名により構成され、「商店テーブル」は「店名」、「Phone.No.」、「所在地」の3つのカラム名により構成されている。各カラムには、夫々、みそラーメン、日立飯店等のカラム値が配置されている。図9は、図7のデータベース例を参照し、図3のPAD図に示す同一見出し生成処理2により生成された単語情報を格納した辞書の例である。D1はNo、D2は見出し文字列、D3は正規語、D4は品詞、D5は分類、D6は所属である。Noは、辞書レコードの番号を示すものであり実施例の説明のために設けた項目である。見出し文字列D2は、入力された文の文字列と比較照合するための文字列であり、正規語D3は、基本的には見出し文字列に対応する文字列であり操作指示対象1であるデータベース内の語を構成する文字列と同一の文字列であり、品詞D4は、語の品詞を示すものである。分類D5は、操作指示対象の世界における語の性格を表わすものであり、図7におけるテーブル名、カラム名、カラム値が入る。所属D6は、操作指示対象1の世界であるデータベース内の位置を示すものであり、図7におけるテーブル名としての具体的な名称(例えば、商品テーブル)またはカラム名としての具体的な名称(例えば、商品名)が入る。また、正規語D3と分類D5と所属D6は操作指示対象の世界に依存する情報であり、見出し文字列D2との対応が取れていれば、辞書3内に設けず、他のテーブルとして設けることも可能である。

6

【0011】以下、同一見出し生成処理を図3のPAD図に従って説明する。まず、操作指示の対象であるデータベースの先頭テーブルから末尾テーブルまで以下の処理を行う(ステップ21)。操作指示対象1のデータベースを参照し、テーブル名(図7の場合、商品テーブル、商店テーブル)を取得し、取得した文字列を見出し文字列として辞書に登録する(ステップ22)。ステップ22により登録した見出し文字列の品詞を「名詞」、分類を「テーブル名」、所属を見出し文字列として辞書に登録する(ステップ23)。ここでいう「品詞」とは単語の品詞を表わし、「分類」とは操作指示対象3のデータベース中の役割を表わし、「所属」とは操作指示対象3のデータベース中の位置を表わしている。また、図3の処理では、見出し文字列の品詞を全て名詞として処理する。次に、操作指示対象1のデータベースを参照し、カラム名(図7の場合、商品名、価格、店名、Phone.No.所在地)を取得し、取得した文字列を見出し文字列として辞書に登録する(ステップ24)。ステップ24により登録した見出し文字列の品詞を「名詞」、分類を「カラム名」、所属を該当するテーブル名として辞書に登録する(ステップ25)。次に、操作指示対象1のデータベースを参照し、カラム値(図7の場合、みそラーメン、ラーメン、……、渋谷、新宿)を取得し、取得した文字列を見出し文字列として辞書に登録する(ステップ26)。ステップ26により登録した見出し文字列の品詞を「名詞」、分類を「カラム値」、所属を該当するカラム名として辞書に登録する(ステップ27)。次に、処理の対象となるテーブルを次のテーブルに移動する(ステップ28)。ステップ21～ステップ28により、図7に示すデータベースの文字列を取得し単語情報を生成し、図9に示すように辞書に格納される。これによりデータベース中の文字列と等しい文字列を見出し文字列とする単語情報が生成される。

【0012】図4は、図1における異表記見出し生成処理の概要を表わすPAD図であり、見出し文字コード変換処理41と同類語見出し生成処理42により構成される。図5は、図4における見出し文字コード変換処理を表わすPAD図である。操作指示対象1であるデータベース内の語を構成する文字列で、他の文字列コードとの対応関係のみにより文字列コードの変換が可能な場合に、見出し文字列を加工し、その加工した文字列を見出し文字列とする単語情報とデータベース内の語との対応関係を持つ単語情報を生成し、辞書3に登録するまでの処理を示したものである。図11は、前記同一見出し生成処理2により生成した図9に示す辞書を参照して、見出し文字コード変換処理41により生成した見出し文字列を持つ単語情報を格納した辞書の例である。D1、D2、D3、D4、D5、D6については図9の場合と同じである。

【0013】以下、この見出し文字コード変換処理をP

7

AD図に従って説明する。まず、前記同一見出し生成処理2により生成した辞書3の先頭レコードから末尾レコードまで以下の処理を行う(ステップ4101)。辞書3の見出し文字列を取得する(ステップ4102)し、この辞書レコードをワーキングエリア内に複写する(ステップ4103)。

【0014】次に、見出し文字列が2バイトコード(実施例の場合、全角文字)のみにより構成されているか否かを判別し(ステップ4104)、構成されている場合

(例えば、図9のNO. 1の「商品テーブル」)に、該2バイトコードに対応する1バイトコード(実施例の場合、半角文字)が存在するか否かを判別し(ステップ4105)、存在する場合(例えば、「商品テーブル」の「テーブル」が半角文字になる)に、該見出し文字列中の1バイトコードに対応する2バイトコードの文字列を1バイトコードに変換し(例えば、「商品テーブル」が図11のNO. 20の「商品」と半角の「テーブル」からなる「商品テーブル」に変換される(なお、JIS変換するときに半角の「テーブル」は変換不可能であるためこのように記載した)) (ステップ4106)、複写した前記辞書レコードの見出し文字列と置換し、辞書3に登録する(ステップ4107)。

【0015】次に、見出し文字列が1バイトコードのみにより構成されているか否かを判別し(ステップ4108)、構成されている場合に、該1バイトコードに対応する2バイトコードが存在するか否かを判別し(ステップ4109)、存在する場合に、該見出し文字列中の2バイトコードに対応する1バイトコードの文字列を2バイトコードに変換し(ステップ4110)、複写した前記辞書レコードの見出し文字列と置換し、辞書3に登録する(ステップ4111)。例えば、図9のNO. 6の「Phone.No」が図11のNO. 23、24の「Phone.No」、「PHONE.NO」に変換される。

【0016】次に、見出し文字列が1バイトコードと2バイトコードにより構成されているか否かを判別し(ステップ4112)、構成されている場合に、該1バイトコードまたは2バイトコードに対応する各々のバイトコードが存在するか否かを判別し(ステップ4113)、存在する場合に、該見出し文字列中の2バイトコードまたは1バイトコードに対応する文字列を各々のバイトコードに変換し(ステップ4114)、複写した前記辞書レコードの見出し文字列と置換し、辞書3に登録する(ステップ4115)。例えば、図9のNO. 12の「A定食」が図11のNO. 39~40の「a定食」、「A定食」、「a定食」に変換される。

【0017】次に、見出し文字列中に平仮名文字列が含まれているか判別し(ステップ4116)、含まれている場合、見出し文字列中の平仮名文字列に対応する片仮名文字列と置換し(ステップ4117)、辞書3に登録する(ステップ4118)。次に、見出し文字列中に片

8

仮名文字列が含まれているか判別し(ステップ4119)、含まれている場合、見出し文字列中の片仮名文字列に対応する平仮名文字列と置換し(ステップ4120)、辞書3に登録する(ステップ4121)。

【0018】ステップ4101~ステップ4121により、図9に示すデータベース中の語を構成する文字列と等しい見出し文字列を持つ単語情報から、図11に示す見出し文字列をコード変換した文字列を見出し文字列とする単語情報を生成し辞書3に格納する。これによりデータベース中の語を構成する文字列のコード変換可能な文字列をコード変換した文字列を見出し文字列とする単語情報が生成される。

【0019】図9のNo.(D1)1の見出し文字列から図11のNo.(D1)19~20の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)2の見出し文字列から図11のNo.(D1)21~22の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)6の見出し文字列から図11のNo.(D1)23~25の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)8の見出し文字列から図11のNo.(D1)26~30の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)9の見出し文字列から図11のNo.(D1)31~32の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)10の見出し文字列から図11のNo.(D1)33~35の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)11の見出し文字列から図11のNo.(D1)36~38の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)12の見出し文字列から図11のNo.(D1)39~41の見出し文字列を持つ辞書レコードが生成される。また、図9のNo.(D1)15の見出し文字列から図11のNo.(D1)42~43の見出し文字列を持つ辞書レコードが生成される。

【0020】図6は、図4における同類語見出し生成処理を表わすPAD図である。操作指示対象1であるデータベース内の語を構成する文字列を利用して同類語(略語)となる見出し文字列を持つ単語情報とデータベース内の語との対応関係を持つ単語情報を生成し、辞書3に登録するまでの処理を示したものである。図8は、図1、図5に示した基本語辞書5の例であり、見出し語文字列B1、品詞B2の2つの項目により構成されている。従来技術の電子化辞書では、この他に文法情報や単語の接続情報等の情報も辞書情報として格納している場合があり、容易に実現可能な技術である。図10は、図8に示す基本語辞書5を参照して、図9に示す単語情報の見出し文字列を形態素解析した結果の例である。W1はNo、W2は見出し文字列、W3は単語分割結果、W4は分割結果別品詞であり、3つの項目W2~W4によ

10

20

30

40

50

り構成されている。単語分割結果W3は見出し文字列の単語分割結果を表わしたものでありカンマ「,」で分割された文字列を区切っている。分割結果別品詞W4は分割された単語別の品詞を表わすものである。図12は、同類語見出し生成処理42のステップ4201～ステップ4207により生成した見出し文字列を持つ単語情報を格納した辞書の例であり、D1、D2、D3、D4、D5、D6については図9の場合と同じである。図13は、同類語見出し生成処理42のステップ4208～ステップ4209により生成した見出し文字列を持つ単語情報を格納した辞書の例であり、D1、D2、D3、D4、D5、D6については図9の場合と同じである。

【0021】以下、この同類語見出し生成処理をPAD図に従って説明する。まず、前記同一見出し生成処理2により生成した辞書3の先頭レコードから末尾レコードまでステップ4202～ステップ4203の処理を行う(ステップ4201)。辞書3の見出し文字列を取得する(ステップ4202)し、基本語辞書5を参照し取得した見出し文字列を形態素解析し、解析結果をワーキングエリアに格納する(ステップ4203)。この形態素解析は既に常用されている従来の機械翻訳処理技術やワードプロセッシング技術により容易に実現可能である。また、ここでいう基本語辞書とは、形態素解析を行うために参照する名詞、固有名詞、動詞、形容詞等の基本的な語を持つ電子化辞書を示している。次に、ワーキングエリアに格納した形態素解析結果の先頭レコードから末尾レコードまで以下の処理を行う(ステップ4204)。まず、形態素解析結果レコードを取得する(ステップ4205)。取得したレコードから単語分割結果(例えば、NO. 1の「テーブル」)を取得し、単語分割結果の中の分割された文字列と等しい文字列を持つ他の形態素解析結果レコード(NO. 2のレコード)を検索し、正規語、品詞、分類、所属を取得する(ステップ4206)。ステップ4206の検索のキーとなった文字列を見出し文字列としてステップ4206で取得した正規語、品詞、分類、所属を与えた単語情報を生成し、辞書3に登録する(ステップ4207)。

【0022】次に、単語分割結果の各文字列(例えば、NO. 1における「商品」、「テーブル」)の先頭文字列を1文字取得し(例えば、NO. 1における「商」、「テ」)、この各1文字を組み合わせた文字列(例えば、「商テ」)を見出し文字として該当する辞書レコードの正規語、品詞、分類、所属を与えた単語情報を生成し、辞書3に登録する(ステップ4208)。次に、単語分割結果の各文字列(例えば、NO. 1における「商品」、「テーブル」)の末尾文字列を1文字取得し(例えば、NO. 1における「品」、「ル」)、この各1文字を組み合わせた文字列(例えば、「品ル」)を見出し文字として該当する辞書レコードの正規語、品詞、分類、所属を与えた単語情報を生成し、辞書3に登録する

(ステップ4208)。

【0023】ステップ4201～ステップ4207により、図10に示すデータベース中の語を構成する文字列と等しい見出し文字列の形態素解析結果から、図12に示す略語を表わす文字列を見出し文字列とする単語情報を生成し辞書3に格納する。図12のNo (D1) 50の見出し文字列(D2)「テーブル」は、図10のNo (D1) 1と2から生成される。また、図12のNo (D1) 51の見出し文字列(D2)「テーブル」は、図10のNo (D1) 1と2から生成される。また、図12のNo (D1) 52の見出し文字列(D2)「ラーメン」は、図10のNo (D1) 8と9から生成される。また、図12のNo (D1) 53の見出し文字列(D2)「ラーメン」は、図10のNo (D1) 8と9から生成される。また、図12のNo (D1) 54の見出し文字列(D2)「日立」は、図10のNo (D1) 13と16から生成される。また、図12のNo (D1) 55の見出し文字列(D2)「日立」は、図10のNo (D1) 13と16から生成される。また、図12のNo (D1) 56の見出し文字列(D2)「鶴亀」は、図10のNo (D1) 14と15から生成される。また、図12のNo (D1) 57の見出し文字列(D2)「鶴亀」は、図10のNo (D1) 14と15から生成される。

【0024】ステップ4208～ステップ4209により、図10に示すデータベース中の語を構成する文字列と等しい見出し文字列の形態素解析結果から、図13に示す略語を表わす文字列を見出し文字列とする単語情報を生成し辞書3に格納する。図13のNo (D1) 60～69は、分割単語の先頭文字の組合せによる略語を表わす見出し文字列を持つ単語情報であり、No (D1) 70～79は、分割単語の末尾文字の組合せによる略語を表わす見出し文字列を持つ単語情報である。また、図13のNo (D1) 60の単語情報「商テ」は、図10のNo (D1) 1から生成される。また、図13のNo (D1) 61の単語情報「商テ」は、図10のNo (D1) 2から生成される。また、図13のNo (D1) 62の単語情報「商名」は、図10のNo (D1) 3から生成される。また、図13のNo (D1) 63の単語情報「P.N」は、図10のNo (D1) 6から生成される。また、図13のNo (D1) 64の単語情報「み」は、図10のNo (D1) 8から生成される。また、図13のNo (D1) 65の単語情報「A定」は、図10のNo (D1) 12から生成される。また、図13のNo (D1) 66の単語情報「日飯」は、図10のNo (D1) 13から生成される。また、図13のNo (D1) 67の単語情報「喫鶴」は、図10のNo (D1) 14から生成される。また、図13のNo (D1) 68の単語情報「か鶴」は、図10のNo (D1) 15から生成される。また、図13のNo (D1) 69の単

11

語情報「日亭」は、図10のNo.(D1)16から生成される。

【0025】また、図13のNo.(D1)70の単語情報「品ル」は、図10のNo.(D1)1から生成される。また、図13のNo.(D1)71の単語情報「品ル」は、図10のNo.(D1)2から生成される。また、図13のNo.(D1)72の単語情報「品名」は、図10のNo.(D1)3から生成される。また、図13のNo.(D1)73の単語情報「e.o」は、図10のNo.(D1)6から生成される。また、図13のNo.(D1)74の単語情報「ソソ」は、図10のNo.(D1)8から生成される。また、図13のNo.(D1)75の単語情報「A食」は、図10のNo.(D1)12から生成される。また、図13のNo.(D1)76の単語情報「立店」は、図10のNo.(D1)13から生成される。また、図13のNo.(D1)77の単語情報「茶亀」は、図10のNo.(D1)14から生成される。また、図13のNo.(D1)78の単語情報「ふえ亀」は、図10のNo.(D1)15から生成される。また、図13のNo.(D1)79の単語情報「立亭」は、図10のNo.(D1)16から生成される。

【0026】図12と図13に示す単語情報を格納した辞書データ例は、操作指示対象であるデータベース内の語を構成する文字列をそのまま見出し文字列として持つ単語情報を基に生成した例であり、図4中の見出し文字コード変換処理により生成された見出し文字列を持つ単語情報(図11)を基に生成すれば、より多種の見出し文字列を持つ単語情報を生成することができる。

【0027】以上、述べたように、本実施例によれば、自然語解析が対象とするデータベースの情報を持つ単語情報を一括して辞書登録することにより、利用者が単語情報を作成し辞書登録する非常に膨大な作業を省くことができるようになる。さらに、データベース内に存在する単語と同一表記(見出し)文字列を持つ単語情報のみではなく、利用者が容易に類推可能であり使用すると予測される同類語や略語等の単語の表記文字列を持つ単語情報を生成し辞書登録することにより、辞書情報を充実することが可能となり、これらの単語の辞書登録についても非常に膨大な作業を省くことができるようになる。

【0028】次に、本発明による辞書生成システムの応用例を説明する。単語情報を生成するデータベースの範囲を利用者に問い合わせ、利用者が指定したテーブル名、カラム名を取得し、単語情報を生成する範囲を限定した後、前記実施例で示した処理を施す。また、見出し文字列の加工(生成)方法を利用者に問い合わせ、利用者が指定した加工方法に合わせて前記実施例で示した処理を施す。拡張した本実施例によれば、自然語解析が対象とするデータベースの持つ単語情報を利用者が指定した範囲のみを効率良く一括して辞書登録することができると同時に、利用者が単語情報を作成し辞書登録する非

12

常に膨大な作業を省くことができるようになる。さらに、データベース内に存在する単語と同一表記文字列を持つ単語情報のみではなく、利用者が容易に類推可能であり使用すると予測される同類語や略語等の単語の表記文字列を持つ単語情報をデータベース内に存在する単語から利用者が指定した加工方法のみにより限定し生成することにより、時間的並びに容量的に効率良く辞書登録し辞書情報を充実することが可能となると同時に、これらの単語の辞書登録についても非常に膨大な作業を省くことができるようになる。

【0029】次に、本発明による辞書生成システムにより生成された辞書の使用例を説明する。図14は、本実施例により生成された辞書を参照する自然語インタフェースの例であり、自然語入力aにより入力された自然語を、本発明で生成された辞書3を参照し操作指示対象命令認識部bにより操作指示命令を認識し、操作指示命令実行部cにより操作指示対象1であるデータベースの操作を行うシステムの機能ブロックである。このように、本発明により生成される辞書を自然語インタフェースで使用することは、容易に類推可能である。

【0030】図15は、図14におけるユーザからの操作指示命令を認識する操作指示命令認識処理bの概要PAD図であり、自然語文の用語を認定する用語認定処理b1、操作指示命令の認識結果を表現する中間表現を生成する中間表現生成処理にb2より構成されている。図16は、図15における用語認定処理b1のPAD図である。以下、この処理をPAD図に従って説明する。ユーザから入力される自然語文字列の入力待ちを行ない(ステップb11)、入力された自然語文字列を文字列変数INPに代入する(ステップb12)。次に、文字列変数INPに文節が含まれている間、以下の処理を行なう(b13)。文節中に辞書3の見出し文字列D2が含まれているか否かを判別し(ステップb14)、含まれている場合に、文節テーブルの語幹M1に見出し文字列D1を格納し、付属語M2にその他の文字列を格納し、品詞M4に品詞D4を格納する(ステップb15)。次に、文節テーブルの正規語M3に正規語D3を格納し、項目分類M5に分類D5を格納し、所属テーブルM6に所属D6を格納する(ステップb16)。次に、文字列変数INPより当該文節文字列分を削除する(ステップb17)。ステップb11～ステップb17により、図18に示す自然語文を入力として用語認定処理b1を行ない、図20に示す文節テーブルに用語認定結果を格納する。ステップb11～ステップb17により、図19に示す自然語文を入力として用語認定処理b1を行ない、図21に示す文節テーブルに用語認定結果を格納する。

【0031】図17は、図15における中間表現生成処理b2のPAD図である。以下、この処理をPAD図に従って説明する。文節テーブルの先頭文節から末尾文節

13

まで、ステップb22～ステップb25の処理を行なう(ステップb21)。項目分類M5を取得し、項目分類M5がカラム値であるか否かを判別し(ステップb22)、カラム値である場合に(b22)、該カラム値に対応するカラム名を取得し(ステップb23)、検索条件としてのカラム名とカラム値の関係を取得する。次に、ステップb21～ステップb25により生成された条件が複数個存在する場合には(ステップb26)、各々の条件を論理積条件として「and」で連結し、これを検索条件として決定する(ステップb27)。生成された条件が複数個存在しない場合には、該当する1つの条件を検索条件として決定する(ステップb28)。次に、文末から文節として1つ以上連続する項目分類がカラム名である文節の正規語を取得しこれを検索対象とする(ステップb29)。ステップb21～ステップb29により、図20に示す文節テーブルを入力として自然語文が表わす操作指示命令を認識し、図22に示す操作指示命令を表現する中間表現を生成する。ステップb21～ステップb29により、図21に示す文節テーブルを入力として自然語文が表す操作指示命令を認識し、図23に示す操作指示命令を表現する中間表現を生成する。

【0032】本実施例では簡単な操作指示命令を認識する処理を示しており、複雑な操作指示命令の認識は、例えば、絹川博之著「表階層モデルに基づく自然語インタフェース処理方式」(情報処理学会論文誌Vol. 27 No. 5 May 1986)のpp. 499～508で示されている従来技術により容易に実現可能である。

【0033】前記中間表現からデータベースの検索言語であるSQL(Structured Query Language)言語へ変換しデータベース検索の実行が可能となる。また、中間表現は形式言語であり、SQL言語への変換は、形式言語から形式言語への変換であり従来技術により容易に実現可能である。

【0034】

【発明の効果】本発明によれば、自然語解析が対象とする操作指示対象の情報を持つ単語情報を一括して辞書登録することにより、利用者が単語情報を作成し辞書登録する非常に膨大な作業を省くことができるようになる。さらに、操作指示対象内に存在する単語と同一表記(見出し)文字列を持つ単語情報のみではなく、利用者が容易に類推可能であり使用すると予測される同類語や略語等の単語の表記文字列を持つ単語情報を生成し辞書登録することにより、辞書情報を充実することが可能となり、これらの単語の辞書登録についても非常に膨大な作業を省くことができるようになるという効果がある。

【図面の簡単な説明】

【図1】本発明の辞書生成システムの一実施例を示す機能ブロック図である。

14

【図2】図1における辞書生成システムの全体的なハードウェア構成を示すブロック図である。

【図3】同一見出し生成処理のPAD図である。

【図4】異表記見出し生成処理の概要PAD図である。

【図5】見出し文字コード変換処理のPAD図である。

【図6】同類語見出し生成処理のPAD図である。

【図7】図1の操作指示対象1であるデータベースの例を示す図である。

【図8】基本語辞書5の例を示す図である。

【図9】図7のデータベース例を参照し、図3のPAD図に示す同一見出し生成処理2により生成された単語情報を格納した辞書の例を示す図である。

【図10】見出し文字列を形態素解析した結果の例を示す図である。

【図11】見出し文字コード変換処理41により生成した見出し文字列を持つ単語情報を格納した辞書の例を示す図である。

【図12】同類語見出し文字生成処理により生成される見出し文字列を持つ単語情報を格納した辞書の例を示す図である。

【図13】同類語見出し文字生成処理により生成される見出し文字列を持つ単語情報を格納した辞書の他の例を示す図である。

【図14】本発明により生成された辞書を参照する自然語インタフェースの例を示す図である。

【図15】図14におけるユーザからの操作指示命令を認識する操作指示命令認識処理bの概要PAD図である。

【図16】図15における用語認定処理b1のPAD図である。

【図17】図15における中間表現生成処理b2のPAD図である。

【図18】操作指示命令を表わす自然語文の例を示す図である。

【図19】操作指示命令を表わす自然語文の他の例を示す図である。

【図20】図18に示す自然語文の用語認定結果を格納した文節テーブルの例を示す図である。

【図21】図19に示す自然語文の用語認定結果を格納した文節テーブルの例を示す図である。

【図22】図20に示す文節テーブルを入力として自然語文が表わす操作指示命令を認識した結果を表現した中間表現の例を示す図である。

【図23】図21に示す文節テーブルを入力として自然語文が表わす操作指示命令を認識した結果を表現した中間表現の例を示す図である。

【符号の説明】

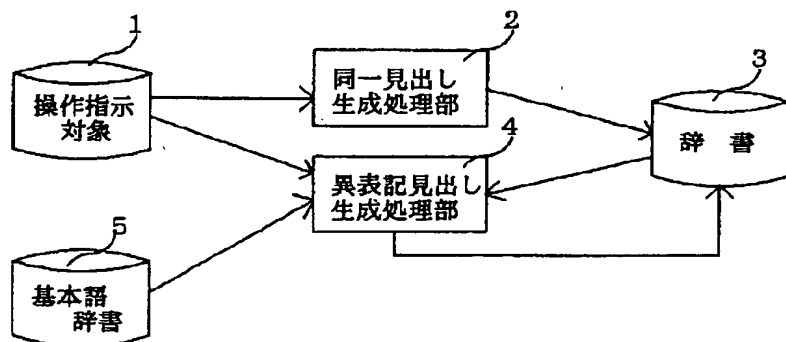
- 1 操作指示対象部
- 2 同一見出し生成処理部
- 3 辞書

- 15
- 4 異表記見出し生成部
5 基本語辞書
6 入出力装置

- 7 プロセッサ
8 記憶装置

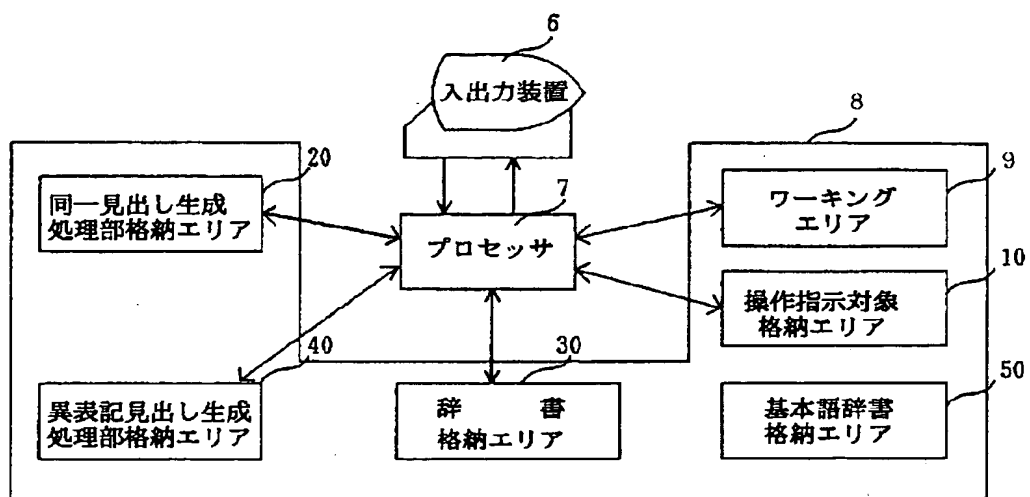
【図1】

【図1】



【図2】

【図2】



【図18】

【図19】

【図18】

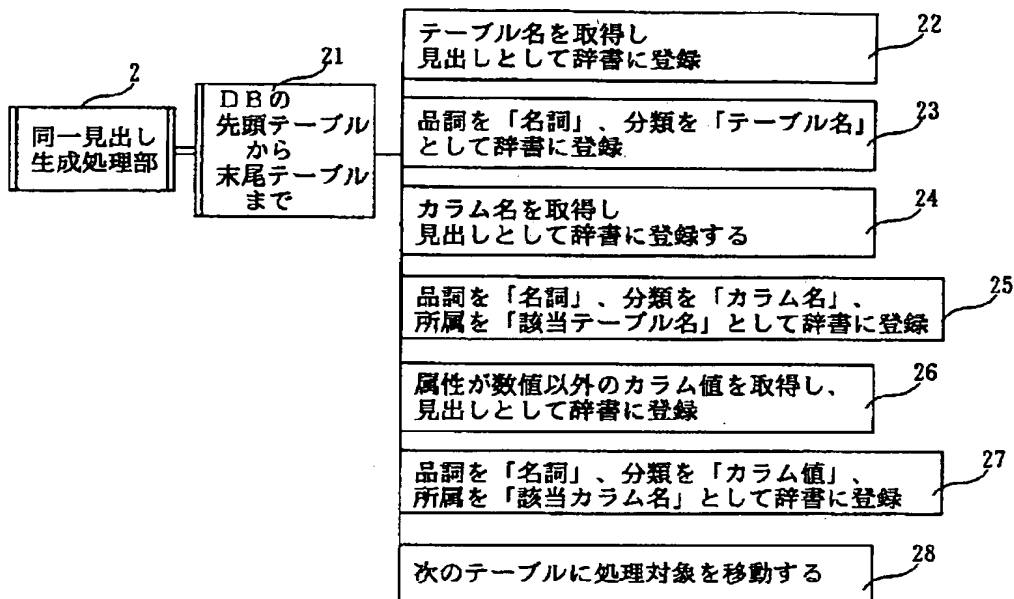
【図19】

ラーメンの価格は

ラーメンの価格は

【図 3】

【図 3】

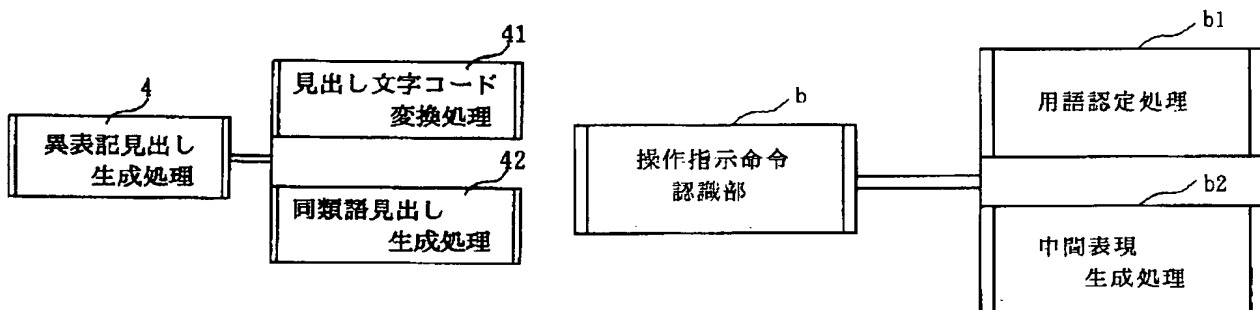


【図 4】

【図 1 5】

【図 4】

【図 1 5】

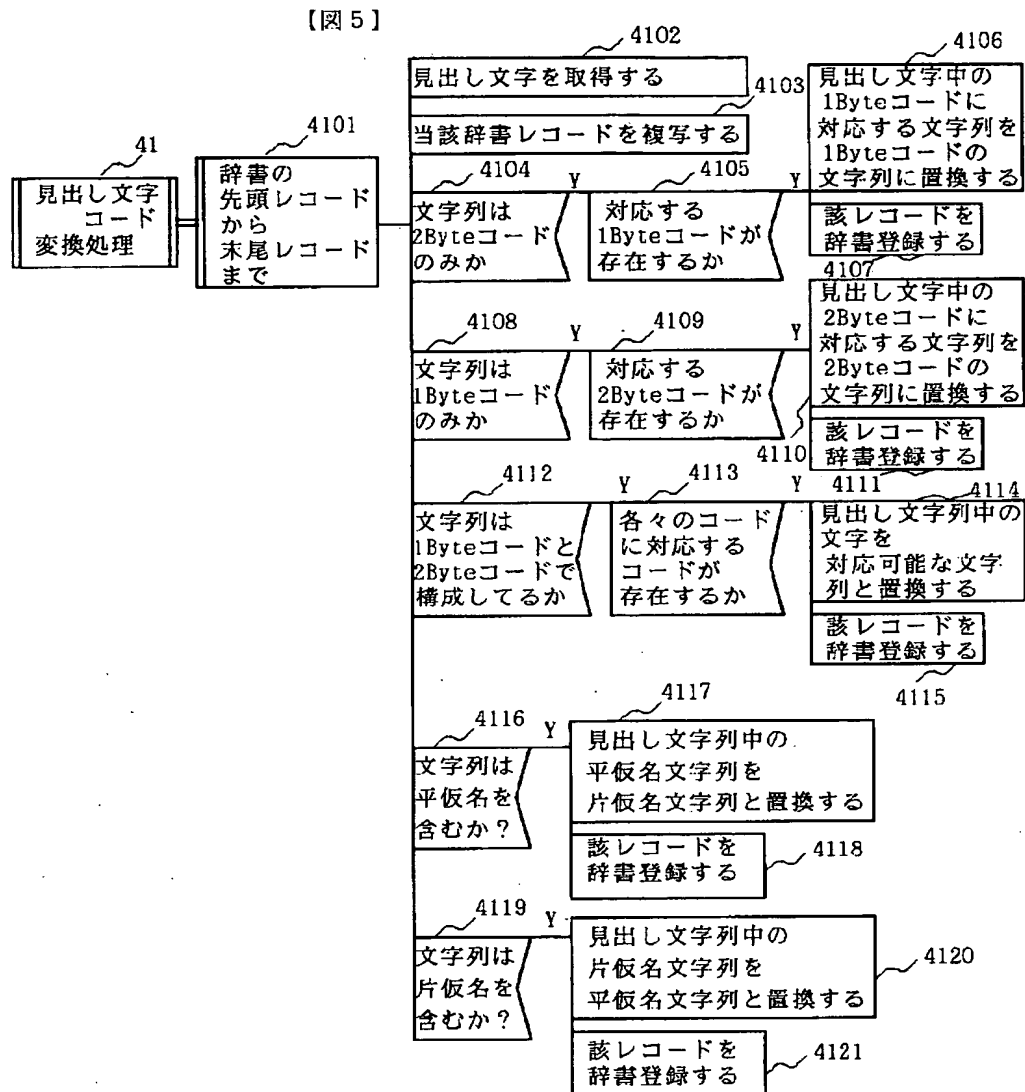


【図 2 0】

【図 2 0】

M 1	M 2	M 3	M 4	M 5	M 6
語 幹	付属語	正規語	品 詞	項目分類	所属テーブル
ラーメン	の	ラーメン みそラーメン	名詞 名詞	カラム値 カラム値	商品名 商品名
価格	は	価格	名詞	カラム名	商品テーブル

【図5】

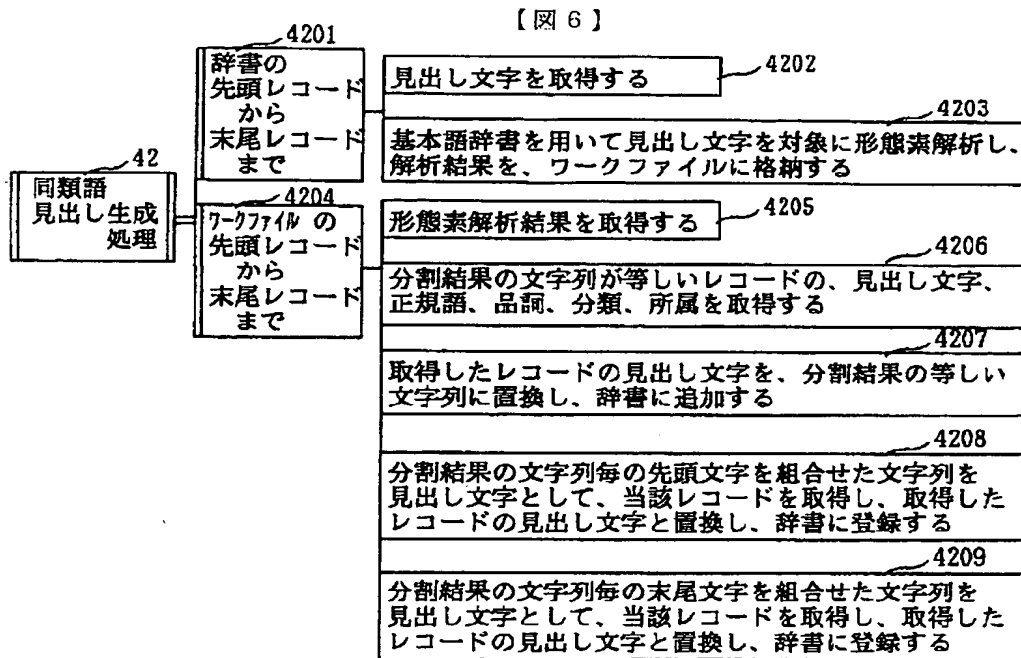


【図21】

【図21】

M 1	M 2	M 3	M 4	M 5	M 6
語 幹	付属語	正規語	品 詞	項目分類	所属テーブル
日立飯店 所在地	の の	日立飯店 所在地	名詞 名詞	コラム値 コラム名	店名 商品テーブル

【図 6】



【図 1 2】

【図 1 2】

D 1	D 2	D 3	D 4	D 5	D 6
N o	見出し文字列	正規語	品詞	分類	所属
50	テーブル	商品テーブル	名詞	テーブル名	商品テーブル
51	テーブル	商店テーブル	名詞	テーブル名	商店テーブル
52	ラーメン	みそラーメン	名詞	カラム値	商品名
53	ラーメン	ラーメン	名詞	カラム値	商品名
54	日立	日立飯店	名詞	カラム値	店名
55	日立	日立亭	名詞	カラム値	店名
56	鶴亀	喫茶鶴亀	名詞	カラム値	店名
57	鶴亀	かふえ鶴亀	名詞	カラム値	店名

【図7】

【図7】

【商品テーブル】

商品名	価格	店名
みそラーメン	500	日立飯店
ラーメン		日立飯店
mixsandwich	450	喫茶鶴亀
Coffee	300	かふえ鶴亀
A定食	1250	日立亭
.	.	.
.	.	.
.	.	.

【商店テーブル】

店名	Phone.No	所在地
日立飯店	256-1234	新宿
喫茶鶴亀	123-4567	渋谷
かふえ鶴亀	321-7654	渋谷
日立亭	248-1632	新宿
.	.	.
.	.	.
.	.	.

【図22】

【図22】

検索条件	商品名 = ラーメン
検索対象	価格

【図8】

【図8】

B 1 見出し文字列	B 2 品詞
商品	名詞
テーブル	名詞
商品	名詞
名	接尾語
価格	名詞
店名	名詞
Phone	名詞
.	記号
No	名詞
所在地	名詞
みそ	名詞
ラーメン	名詞
mixsandwich	名詞
Coffee	名詞
A	記号
定食	名詞
日立	名詞
飯店	名詞
喫茶	名詞
鶴亀	名詞
かふえ	名詞
亭	接尾語
新宿	名詞
渋谷	名詞

【図23】

【図23】

検索条件	商店名 = 日立飯店
検索対象	所在地

【図9】

【図9】

D 1	D 2	D 3	D 4	D 5	D 6
N o	見出し文字列	正規語	品詞	分類	所属
1	商品テーブル	商品テーブル	名詞	テーブル名	商品テーブル
2	商店テーブル	商店テーブル	名詞	テーブル名	商店テーブル
3	商品名	商品名	名詞	カラム名	商品テーブル
4	価格	価格	名詞	カラム名	商品テーブル
5	店名	店名	名詞	カラム名	商品テーブル, 商店テーブル
6	Phone.No	Phone.No	名詞	カラム名	商店テーブル
7	所在地	所在地	名詞	カラム名	商店テーブル
8	みそラーメン	みそラーメン	名詞	カラム値	商品名
9	ラーメン	ラーメン	名詞	カラム値	商品名
10	mixsandwitch	mixsandwitch	名詞	カラム値	商品名
11	C o f f e e	C o f f e e	名詞	カラム値	商品名
12	A定食	A定食	名詞	カラム値	商品名
13	日立飯店	日立飯店	名詞	カラム値	店名
14	喫茶鶴亀	喫茶鶴亀	名詞	カラム値	店名
15	かふえ鶴亀	かふえ鶴亀	名詞	カラム値	店名
16	日立亭	日立亭	名詞	カラム値	店名
17	新宿	新宿	名詞	カラム値	所在地
18	渋谷	渋谷	名詞	カラム値	所在地

【図10】

【図10】

W1	W2	W3	W4
N o	見出し文字列	単語分割結果	分割結果別品詞
1	商品テーブル	商品, テーブル	名詞, 名詞
2	商店テーブル	商店, テーブル	名詞, 名詞
3	商品名	商品, 名	名詞, 接尾語
4	価格	価格	名詞
5	店名	店名	名詞
6	Phone.No	Phone, ., No	名詞, 記号, 名詞
7	所在地	所在地	名詞
8	みそラーメン	みそ, ラーメン	名詞, 名詞
9	ラーメン	ラーメン	名詞
10	mixsandwitch	mixsandwitch	名詞
11	C o f f e e	C o f f e e	名詞
12	A定食	A, 定食	記号, 名詞
13	日立飯店	日立, 飯店	名詞, 名詞
14	喫茶鶴亀	喫茶, 鶴亀	名詞, 名詞
15	かふえ鶴亀	かふえ, 鶴亀	名詞, 名詞
16	日立亭	日立, 亭	名詞, 接尾語
17	新宿	新宿	名詞
18	渋谷	渋谷	名詞

【図11】

【図11】

D1	D2	D3	D4	D5	D6
No	見出し文字列	正規語	品詞	分類	所属
19	商品てーぶる	商品テーブル	名詞	テーブル名	商品テーブル
20	商品てーブル	商品テーブル	名詞	テーブル名	商品テーブル
21	商店てーぶる	商店テーブル	名詞	テーブル名	商店テーブル
22	商店てーブル	商店テーブル	名詞	テーブル名	商店テーブル
23	Phone.No	Phone.No	名詞	カラム名	商店テーブル
24	PHONE.NO	Phone.No	名詞	カラム名	商店テーブル
25	phone.no	Phone.No	名詞	カラム名	商店テーブル
26	ミソラーメン	みそラーメン	名詞	カラム値	商品名
27	ミソらーめん	みそラーメン	名詞	カラム値	商品名
28	ミソら	みそラーメン	名詞	カラム値	商品名
29	ミソらーめん	みそラーメン	名詞	カラム値	商品名
30	みそらーめん	みそラーメン	名詞	カラム値	商品名
31	らーめん	ラーメン	名詞	カラム値	商品名
32	ら	ラーメン	名詞	カラム値	商品名
33	mixsandwich	mixsandwich	名詞	カラム値	商品名
34	MIXSANDWICH	mixsandwich	名詞	カラム値	商品名
35	MIXSANDWICH	mixsandwich	名詞	カラム値	商品名
36	COFFEE	Coffee	名詞	カラム値	商品名
37	Coffee	Coffee	名詞	カラム値	商品名
38	COFFEE	Coffee	名詞	カラム値	商品名
39	a定食	A定食	名詞	カラム値	商品名
40	A定食	A定食	名詞	カラム値	商品名
41	a定食	A定食	名詞	カラム値	商品名
42	カフェ鰻	かふえ鰻	名詞	カラム値	店名
43	カフェ鰻	かふえ鰻	名詞	カラム値	店名

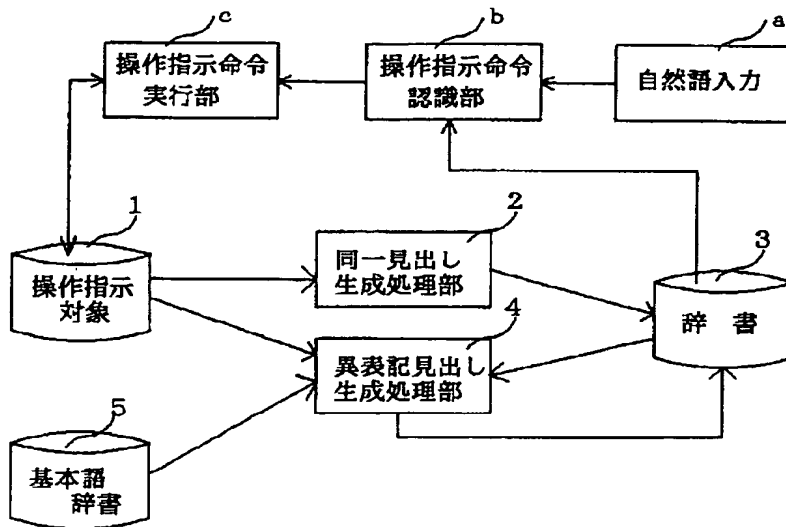
【図13】

【図13】

D 1	D 2	D 3	D 4	D 5	D 6
N o	見出し文字列	正規語	品詞	分類	所属
60	商テ	商品テーブル	名詞	テーブル名	商品テーブル
61	商テ	商店テーブル	名詞	テーブル名	商店テーブル
62	商名	商品名	名詞	カラム名	商品テーブル
63	P.N	Phone.No	名詞	カラム名	商品テーブル
64	みラ	みそラーメン	名詞	カラム値	商品名
65	A定	A定食	名詞	カラム値	商品名
66	日飯	日立飯店	名詞	カラム名	店名
67	喫鶴	喫茶鶴亀	名詞	カラム名	店名
68	か鶴	かふゑ鶴亀	名詞	カラム名	店名
69	日亭	日立亭	名詞	カラム名	店名
70	品ル	商品テーブル	名詞	テーブル名	商品テーブル
71	品ル	商店テーブル	名詞	テーブル名	商店テーブル
72	品名	商品名	名詞	カラム名	商品テーブル
73	e.o	Phone.No	名詞	カラム名	商品テーブル
74	そン	みそラーメン	名詞	カラム値	商品名
75	A食	A定食	名詞	カラム値	商品名
76	立店	日立飯店	名詞	カラム名	店名
77	茶亀	喫茶鶴亀	名詞	カラム名	店名
78	ふゑ亀	かふゑ鶴亀	名詞	カラム名	店名
79	立亭	日立亭	名詞	カラム名	店名

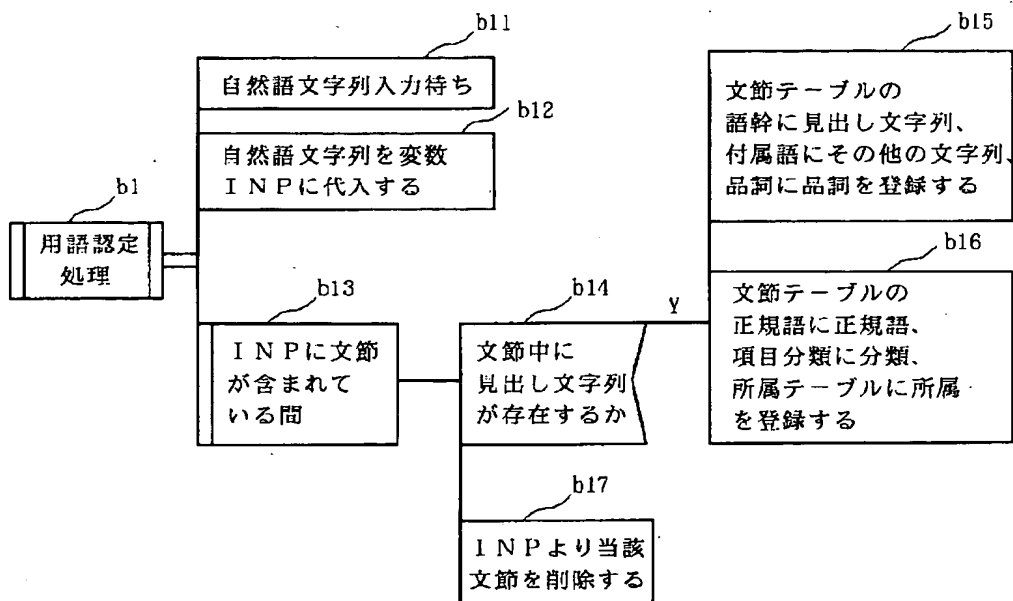
【図 1 4】

【図 1 4】



【図 1 6】

【図 1 6】



【図 1 7】

【図 1 7】

